

# Kmer mapping

BY Bernadette C Young

Updated date: Mar 7, 2021

An abbreviated version of this protocol was published in eLIFE in Feb 2019

Panton–Valentine leucocidin is the key determinant of *Staphylococcus aureus* pyomyositis in a bacterial GWAS

DOI: 10.7554/eLife.42486

## Detailed protocol

We used Bowtie (Langmead and Salzberg, 2012) to align all 31 bp kmers from short-read sequencing to a draft reference (the de novo assembly of a CC-121 pyomyositis strain PYO2134).

Step 1. Make a bowtie reference from the preferred assembly

Example command line

```
> [local installation of bowtie2-build] -f [contigs.ref.fa] [ref_name]
```

Step 2. Align kmers to reference (can be done for all kmers, or a significant subset). This reports only one random chosen mapping if multiple mappings exist

Example command line

```
> bowtie2 -r -D 24 -R 3 -N 0 -L 18 -i S,1,0.30 -x [ref_name] -U [kmer_file_name.txt]
```

The output file gives a position within the reference genome of each kmer in the input, along with mapping quality, which can be used to identify the likely location of each kmer.

Areas of homology between the draft reference and well-annotated reference strains were identified by aligning sequences with Mauve (Darling et al., 2004).

This was performed using the graphic user interface of the Mauve software. Further instructions for the use of Mauve are found at the group website.

(<http://darlinglab.org/mauve/mauve.html>)

The backbone output file of alignment gives the areas of homology between two sequences (in this case a fully annotated reference USA300\_FPR3757 with the de novo assembly of a CC-121 pyomyositis strain PYO2134)

For all 31 bp kmers with significant association with case-controls status, the likely origin of the kmer was determined by nucleotide sequence BLAST (Altschul et al., 1990) of the kmers against a database of all *S.aureus* sequences in Genbank.

This step uses the LMM\_kmer\_annotation script in the Bugwas kmer gwas pipeline which can be found here

<https://github.com/jessiewu/bacterialGWAS>

There is a manual for use

[https://github.com/janepipistrelle/bacterial\\_GWAS\\_tutorial/blob/master/tutorial.rmd](https://github.com/janepipistrelle/bacterial_GWAS_tutorial/blob/master/tutorial.rmd)

**How to cite:** (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Young, B. (2021). Kmer mapping. Bio-protocol Preprint. [bio-protocol.org/1000000](https://doi.org/10.21956/bio-protocol.d1000000).
2. Young, B. C., Earle, S. G., Soeng, S., Sar, P., Kumar, V., Hor, S., Sar, V., Bousfield, R., Sanderson, N. D., Barker, L., Stoesser, N., Emary, K. R., Parry, C. M., Nickerson, E. K., Turner, P., Bowden, R., Crook, D. W., Wyllie, D. H., Day, N. P., Wilson, D. J. and Moore, C. E. (2019). Panton–Valentine leucocidin is the key determinant of *Staphylococcus aureus* pyomyositis in a bacterial GWAS. eLIFE. DOI: [10.7554/eLife.42486](https://doi.org/10.7554/eLife.42486)

**Copyright:** Content may be subjected to copyright.